Supplemental Information

The origin of consistent protein structure refinement from structural averaging

Hahnbeom Park, Frank DiMaio, and David Baker

Inventory of Supplemental Information

1. Supplementary Figure

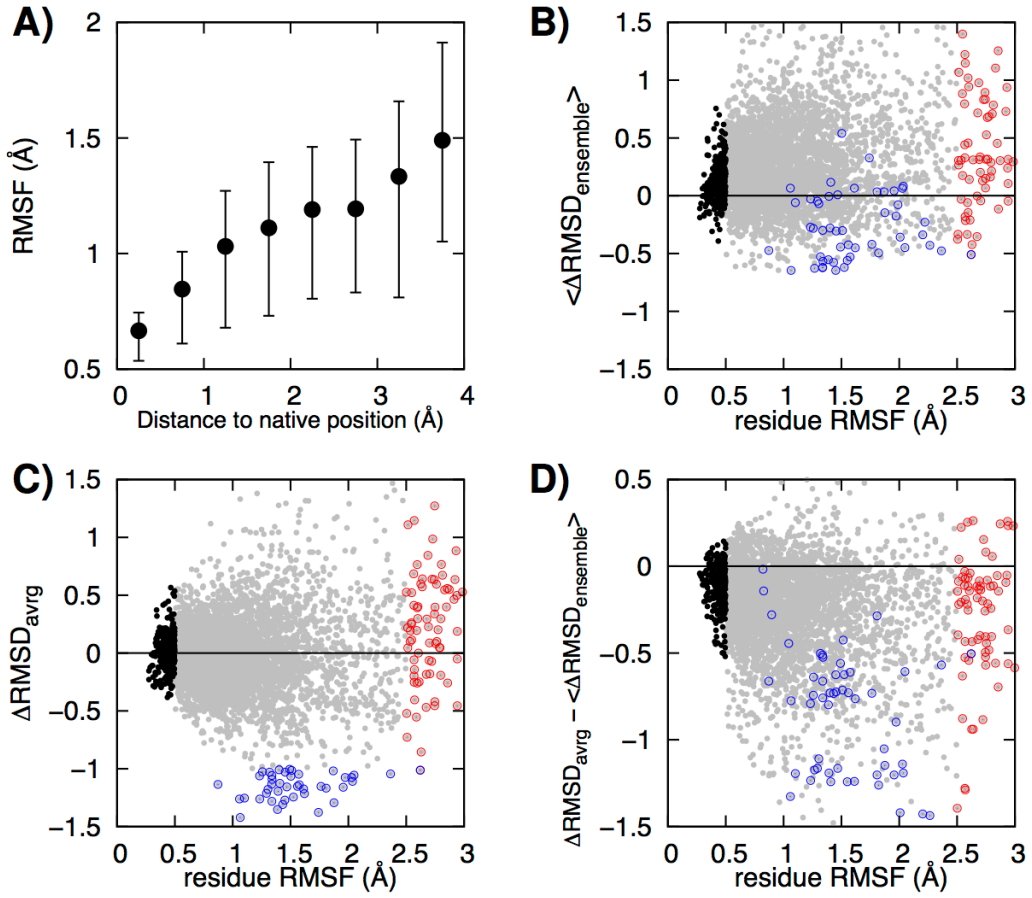

Figure S1. The effect of averaging, for the unrestrained simulations, on per-residue accuracy as function of the magnitude of fluctuation. See **Figure 2** for the results on restrained simulations. **A)** The magnitude of the fluctuations in the coordinates of a residue during the trajectories increases with increasing distance of the starting coordinates from the native structure. The first and third quartiles are shown as error bars. From **B**) to **D**), residue model quality changes are plotted (Y-axis) as a function of the fluctuations of the residue during the simulation (X-axis, root-mean-square fluctuation in Å) for the 40 targets in the study. **B**) The mean value of the per-residue change in RMSD (ΔRMSD) for each member of the ensemble. ΔRMSD values are computed for each residue in each member of the ensemble, and the resulting ΔRMSD values are then averaged. **C**) The per-residue ΔRMSD of the ensemble averaged structure. Members of the ensemble are first structurally averaged, and then the per-residue ΔRMSD is computed for this averaged structure. **D**) The difference between the pre (panel B) and post (panel C) structural averaging per-residue ΔRMSDs**.** The Y-axis indicates the contribution of structural averaging: negative values represent improvements upon averaging. Residues with low fluctuation (< 0.5 Å) or high fluctuation (>2.5 Å) or with large improvements (< -1.0 Å) in the averaged structures are indicated with black, red and blue circles respectively. Residue-level model quality is measured by residue RMSD on a 9-residue window with the target residue at the center.

2. Supplementary Tables

Supplementary Table S1. Refinement in stereochemistry measured by Gaia (Kota, Ding, Ramachandran, Dokholyan, 2011). Definitions of the metrics are described in Kota et al. Closer to the optimal values mean better stereochemistry. Average values and standard deviations over 40 targets are reported.

	Optimal value	Starting structures	Refined structures
clash score	0.015	0.075 (±0.070)	0.009 (±0.007)
unsatisfied Hbonds in shell (%)	6.2	11.1 (± 3.9)	9.1 (± 2.6)
unsatisfied Hbonds in core (%)	0.01	0.67 (± 0.67)	0.60 (± 0.57)
normalized molecular surface area ($Å^2$)	125	128 (± 11.6)	131 (± 10.5)
void volume score	0.2	0.98 (± 0.59)	0.91 (± 0.59)

3. Method availability

Rosetta *CartesianRefiner* is available in the current release of Rosetta software suite weekly update (since Jan 10th, 2015). *CartesianRefiner* should be compiled with MPI option – please refer to Rosetta building guide at:
https://www.rosettacommons.org/docs/latest/Build-Documentation.html.

4. Running scripts

The only input required for the protocol is input structure. To run the protocol using MPI (in linux system),
mpirun <rosetta_source_dir>/bin/mpi_refinement.mpi.linuxgccrelease <flags>

Typical flags are required:

flags:
-database <rosetta_database_path>
-in:file:s <input pdb>
-lh:mpi_master_schfile run.cmd
-mute all
-unmute MPI.LHR
-wum:n_masters 1
-relax:constrain_relax_to_start_coords
-out:file:silent_struct_type binary
-lh:max_emperor_lib_size 40
-lh:max_ref_lib_size 1
-lh:max_lib_size 200
-lh:mpi_feedback add_n_limit
-in:file:native <native pdb> (optional)

with a command file (run.cmd in the flags):

rank 0
iter 1
round search
exclude MDfacts MDfactsSoft

# name	mover	nrun	nper	score	cstw	relaxtype	rerelax	shave1	shave2	shave3
MDfacts	md	5	20	scorefacts	1.00	1	12	0.5	0.5	0.8
MDfactsSoft	md	5	20	scorefacts	1.00	2	12	0.5	0.5	0.8
MD150K	md	12	20	talaris2013	0.25	1	12	0.0	0.9	0.8
MDSoft	md	12	20	talaris2013	0.25	2	12	0.0	0.9	0.8
RlxCartfacts	relax	25	1	scorefacts	1.00	3	12	0.0	0.0	0.8
RlxDualSoft	relax	5	5	talaris2013	0.25	4	11	0.0	0.0	0.5
RlxCart	relax	5	5	talaris2013	0.25	1	11	0.0	0.0	0.5

round wait

round enrich
round wait
round calcdev #not only avrg but also report rmsf
done